



Short and squeezed: accelerating the computation of antispase representations with safe squeezing

Clément Elvira, Cédric Herzet

► To cite this version:

Clément Elvira, Cédric Herzet. Short and squeezed: accelerating the computation of antispase representations with safe squeezing. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2020, Barcelona, Spain. pp.5615-5619, 10.1109/ICASSP40776.2020.9053156 . hal-03070344

HAL Id: hal-03070344

<https://hal.inria.fr/hal-03070344>

Submitted on 15 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SHORT AND SQUEEZED: ACCELERATING THE COMPUTATION OF ANTISPARSE REPRESENTATIONS WITH SAFE SQUEEZING

Clément Elvira and Cédric Herzet

INRIA Rennes-Bretagne Atlantique, Campus de Beaulieu, 35000 Rennes, France
 firstname.lastname@inria.fr

ABSTRACT

Antisparsity coding aims at spreading the information uniformly over representation coefficients and can be expressed as the solution of an ℓ_∞ -norm regularized problem. In this paper, we propose a new methodology, coined “safe squeezing”, accelerating the computation of antisparsity representations. The idea consists in identifying saturated entries of the solution via simple tests and compact their contribution to achieve some form of dimensionality reduction. Numerical experiments show that the proposed approach leads to significant computational gain.

Index Terms— antisparsity coding, safe squeezing, scaled projected gradient, convex optimization

1. INTRODUCTION

Over the past decades, convexity has proven to be a powerful tool for regularizing ill-posed inverse problems. For example, the “ ℓ_1 -norm” penalty has been widely advocated for promoting sparsity in the solution of an optimization problem. When combined with a quadratic discrepancy, such a regularizing function leads to the well-known LASSO problem. Two elements grounding the success of the LASSO are the existence of guarantees ensuring the robust identification of sparse vectors (see [1]) and the advent of numerical procedures able to solve efficiently large-scale problems *e.g.*, [2].

Of particular interest in this paper is an acceleration method first proposed by El Ghaoui *et al.* in the context of sparsity-promoting convex problems, namely “safe screening” [3]. This acceleration procedure revolves around two ingredients. First, the solutions of such optimization problems are known to contain many zeros, see *e.g.*, [1, Th 3.1] for the LASSO. Second, if the positions of (some of) the zeros in the solution are known, the optimization problem can be transformed into a problem of smaller dimension, potentially resulting in huge memory and computational savings.

Over the past few years, safe screening has sparked a surge of interest in the LASSO literature *e.g.*, [4–7] and beyond ℓ_1 regularization [8–10]. However, up to our knowl-

edge, all these contributions exclusively focus on the resolution of “sparsity-promoting” problems. In this paper, we show that the principles ruling safe screening can be extended to a family of convex optimization problems of different nature, namely *antisparsity coding*. More particularly, we address the following problem:

$$\mathbf{x}^* \in \arg \min_{\mathbf{x} \in \mathbf{R}^n} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_\infty, \quad (1)$$

where $\mathbf{y} \in \mathbf{R}^m$ is an observation vector, $\mathbf{A} \in \mathbf{R}^{m \times n}$ a representation matrix and $\lambda > 0$ a penalization parameter. In (1) the admissible range of the coefficients \mathbf{x} is penalized through an ℓ_∞ -norm. For this reason, the solutions of (1) are sometimes referred to as “antisparsity” or “spread” since, contrary to the LASSO problem, they are known to be dense with many “saturated” entries satisfying

$$|\mathbf{x}^*(i)| = \|\mathbf{x}^*\|_\infty. \quad (2)$$

Spreading the information uniformly over the coefficients of a representation is a desirable property in various applications, see *e.g.*, [11–16].

From a numerical point of view, although generic convex optimization tools may be applied to problem (1), the design of algorithms specifically dedicated to this problem has received much less attention than its ℓ_1 counterpart. In this contribution, we show that the main principles underlying “safe screening” may be extended to problem (1). To be more precise, we first show that (1) can be reduced into a problem of smaller dimension when some “saturated” components of the solution \mathbf{x}^* are identified. We then propose a new methodology, coined “safe squeezing”, allowing to uncover some of the saturated components of the solution

The rest of the paper is organized as follows. Section 2 describes the proposed safe squeezing methodology and the problem’s dimensionality reduction it induces. In Section 3 we discuss an algorithmic solution to solve the equivalent “reduced” optimization problem. In Section 4, we illustrate the performance of the proposed methodology in numerical experiments. Concluding remarks are made in Section 5. For sake of brevity, the proofs of our results are omitted in this paper but can be found in [17].

Part of this work has been funded thanks to the Becose ANR project no. ANR-15-CE23-0021.

2. SAFE SQUEEZING

Prior to presenting our proposed safe squeezing methodology, we state two working hypotheses that are assumed to be met along the paper. First, we assume that

$$0 < \lambda < \|\mathbf{A}^T \mathbf{y}\|_1. \quad (3)$$

The left inequality prevents from removing the penalization of problem (1); the right inequality ensures that the solution of problem (1) is not the all-zero vector $\mathbf{0}_n$. For future use, we introduce the notation $\lambda_{\max} \triangleq \|\mathbf{A}^T \mathbf{y}\|_1$. Second, we assume that (1) admits one unique minimizer \mathbf{x}^* . The reader is referred to [17] for a discussion on the case where the latter condition is not met.

2.1. Dimensionality reduction via saturation detection

In this section, we illustrate how the knowledge of the positions of saturated entries in \mathbf{x}^* can lead to memory and complexity savings in the resolution of (1). Let

$$\mathcal{I}_+^* \triangleq \{i \mid \mathbf{x}^*(i) = +\|\mathbf{x}^*\|_\infty\}, \quad (4a)$$

$$\mathcal{I}_-^* \triangleq \{i \mid \mathbf{x}^*(i) = -\|\mathbf{x}^*\|_\infty\}, \quad (4b)$$

be the sets of positive and negative saturated components of \mathbf{x}^* , and $\mathcal{I}^* = \mathcal{I}_+^* \cup \mathcal{I}_-^*$. Then, for any $\mathcal{I} = \mathcal{I}_+ \cup \mathcal{I}_-$ with $\mathcal{I}_+ \subseteq \mathcal{I}_+^*$ and $\mathcal{I}_- \subseteq \mathcal{I}_-^*$, problem (1) can equivalently¹ be rewritten as

$$\begin{aligned} (w^*, \mathbf{q}^*) \in & \arg \min_{(w, \mathbf{q}) \in \mathbf{R} \times \mathbf{R}^{n - \text{card}(\mathcal{I})}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}_{\bar{\mathcal{I}}} \mathbf{q} - \mathbf{s} w\|_2^2 + \lambda w \\ & \text{subject to } \mathbf{q} \leq w, -\mathbf{q} \leq w \end{aligned} \quad (5)$$

where $\mathbf{A}_{\bar{\mathcal{I}}}$ denotes the submatrix of \mathbf{A} whose columns indexed by \mathcal{I} have been removed, “ $\mathbf{q} \leq w$ ” means “ $q(i) \leq w$ for all i ” and

$$\mathbf{s} \triangleq \sum_{i \in \mathcal{I}_+} \mathbf{a}_i - \sum_{i \in \mathcal{I}_-} \mathbf{a}_i. \quad (6)$$

Problem (5) has to be understood as follows: once the position and the sign of (some of) the saturated entries are known, their contribution in the reconstruction of \mathbf{y} can be compacted into a single vector. This is the meaning of vector \mathbf{s} in (6), and w is meant to be the absolute value of the saturation. Then, the two constraints $\pm \mathbf{q} \leq w$ ensures that the absolute value of the coefficients weighting the elements of $\mathbf{A}_{\bar{\mathcal{I}}}$ is no larger than w .

Note that although the two problems are equivalent, their dimensionality can be quite different: whereas the initial problem manipulates n -dimensional variables, problem (5) only involves $n - \text{card}(\mathcal{I}) + 1$ variables. We are particularly interested in the limit case when $\mathcal{I} = \mathcal{I}^*$ where the dimensionality reduction may be quantified by the following lemma [17, Lemma 1]:

¹that is, for all subset $\mathcal{I} \subseteq \mathcal{I}^*$, there exists a bijection between the solution of (5) and the solution of (1).

Lemma 1. *If $\text{kruskal}(\mathbf{A}) = m$ then $\text{card}(\mathcal{I}^*) \geq n - m + 1$.*

Here, $\text{kruskal}(\mathbf{A})$ denotes the Kruskal rank of \mathbf{A} and corresponds to the smallest number of columns of \mathbf{A} that are linearly dependent. As a corollary of Lemma 1, when $\mathcal{I} = \mathcal{I}^*$, the size of problem (5) potentially drops down to m , resulting in a dramatic dimensionality reduction. Hence, addressing (5) instead of (1) is obviously of interest to save computational resources. However, the relevance of problem (5) is conditioned to the identification of a (large) subset of \mathcal{I}^* . This problem is addressed in the next section.

2.2. Safe squeezing test

In this section, we present our proposed *safe squeezing* procedure. It consists in a test ensuring that an index i belongs to \mathcal{I}^* , and relies on the following result [17, Theorem 1]:

Theorem 1. *Let $\mathcal{I} \subseteq \mathcal{I}^*$ and*

$$\mathbf{u}^* = \arg \max_{\mathbf{u} \in \mathcal{U}_{\mathcal{I}}} \frac{1}{2} \|\mathbf{y}\|_2^2 - \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_2^2 \quad (7)$$

where

$$\mathcal{U}_{\mathcal{I}} \triangleq \{\mathbf{u} : \|\mathbf{A}_{\mathcal{I}}^T \mathbf{u}\|_1 + \mathbf{s}^T \mathbf{u} \leq \lambda\} \quad (8)$$

and \mathbf{s} is defined as in (6). Then,

$$\forall i : |\mathbf{a}_i^T \mathbf{u}^*| > 0 \Rightarrow i \in \mathcal{I}_{\text{sign}(\mathbf{a}_i^T \mathbf{u}^*)}^* \quad (9)$$

In other words, the left-hand-side of (9) ensures that the i -th entry of \mathbf{x}^* — the solution of (1) — belongs to \mathcal{I}^* as soon as the i -th column of \mathbf{A} is not orthogonal to some vector \mathbf{u}^* defined in (7). We emphasize that when some i passes the test, the sign of the saturation is also an output of the test. A proof of this result can be found in [17, App. A]. In a few words, (9) results from a complementary slackness condition. We also mention that (7) corresponds to the dual problem of (5) while (8) defines the dual feasible set. For this reason, any vector $\mathbf{u} \in \mathcal{U}_{\mathcal{I}}$ will be referred to as *dual feasible*.

Unfortunately, in practice, identifying a maximizer of (7) turns out to be as difficult as solving the initial optimization problem (5). To overcome this issue, we now devise a relaxed version of the test presented in (9) to identify subsets of \mathcal{I}^* with a low computational burden. Given some $\mathbf{c} \in \mathbf{R}^m$ and $r > 0$, let

$$\mathcal{B}(\mathbf{c}, r) \triangleq \{\mathbf{u} \in \mathbf{R}^m \mid \|\mathbf{u} - \mathbf{c}\|_2 \leq r\} \quad (10)$$

be some spherical region such that

$$\mathbf{u}^* \in \mathcal{B}(\mathbf{c}, r). \quad (11)$$

In the screening literature, such a subset is usually referred to as “safe sphere” and the same convention is used hereafter. Then, the following result holds true (see [17] for the proof)

Theorem 2 (Safe sphere squeezing test). *If $\mathbf{u}^* \in \mathcal{B}(\mathbf{c}, r)$, then*

$$|\mathbf{a}_i^T \mathbf{c}| > r \|\mathbf{a}_i\|_2 \Rightarrow i \in \mathcal{I}_{\text{sign}(\mathbf{a}_i^T \mathbf{c})}^* \quad (12)$$

In other words, the squeezing test (12) provides a practical way of testing whether $i \in \mathcal{I}^*$ since it only requires to evaluate a single inner product (as long as a safe sphere $\mathcal{B}(\mathbf{c}, r)$ is available). However, we emphasize that passing (12) is only a *sufficient* condition for $i \in \mathcal{I}^*$. Obviously, the performance of the safe sphere squeezing test depend on the choice of the center and the radius of the safe sphere.

2.3. Construction of the GAP safe spheres

In this section, we elaborate on the construction of safe spheres, *i.e.*, regions \mathcal{B} verifying (11). In this task, we can benefit from the recent developments in the screening literature since several safe spheres derived in the context of safe screening for LASSO can be reused for safe squeezing up to minor modifications. The rationale behind this result is that the LASSO dual problem shares the same cost function as problem (7) but with a definition of the dual feasible set different from (8), see *e.g.*, [6, Eq. (2)]. In addition, many safe spheres proposed in the context of safe screening does not rely on the specific definition of the dual feasible set.

In this work, we focus on the ‘‘GAP’’ sphere proposed in [4], which can be extended as follows in the context of safe squeezing: for any $\mathcal{I} \subseteq \mathcal{I}^*$ and $(w, \mathbf{q}, \mathbf{u})$ verifying

$$\mathbf{q} \leq w, \quad -\mathbf{q} \leq w, \quad \mathbf{u} \in \mathcal{U}_{\mathcal{I}}, \quad (13)$$

where $\mathcal{U}_{\mathcal{I}}$ has been defined in (8), the sphere $\mathcal{B}(\mathbf{c}, r)$ with

$$\begin{aligned} \mathbf{c} &= \mathbf{u} \\ r &= \sqrt{2 \text{gap}(w, \mathbf{q}, \mathbf{u})} \end{aligned} \quad (14)$$

and

$$\begin{aligned} \text{gap}(w, \mathbf{q}, \mathbf{u}) &= \frac{1}{2} \|\mathbf{y} - \mathbf{A}_{\mathcal{I}} \mathbf{q} - \mathbf{s} w\|_2^2 + \lambda w \\ &\quad - \left(\frac{1}{2} \|\mathbf{y}\|_2^2 - \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_2^2 \right) \end{aligned} \quad (15)$$

is safe for problem (7). The interest of the GAP sphere is twofold. First, using [18, Proposition 5.2.1], one can show that there is no duality gap between (5) and (7). Hence, particularizing the GAP sphere to $w^*, \mathbf{q}^*, \mathbf{u}^*$, one obtains $r = 0$ and the GAP sphere squeezing test (12) boils down to (9). Second, the duality gap is a standard stopping criterion in optimization. In this context, computing the radius of the GAP sphere can thus be done at *no* extra computational cost.

The last computational difficulty in evaluating (14) stands in the identification of a dual feasible point (that is $\mathbf{u} \in \mathcal{U}_{\mathcal{I}}$). However, the ‘‘dual scaling’’ procedure proposed in [3, Section 3.3] can be extended to the present setup. More precisely, given any $\mathbf{z} \in \mathbf{R}^m$, straightforward calculations show that

$$\mathbf{u} = \begin{cases} \mathbf{z} & \text{if } \|\mathbf{A}_{\mathcal{I}}^T \mathbf{z}\|_1 + \mathbf{s}^T \mathbf{z} \leq 0, \\ \frac{\lambda}{\|\mathbf{A}_{\mathcal{I}}^T \mathbf{z}\|_1 + \mathbf{s}^T \mathbf{z}} \mathbf{z} & \text{otherwise.} \end{cases} \quad (16)$$

is dual feasible. Note that (16) only requires the computation of $\text{card}(\bar{\mathcal{I}}) + 1$ inner products in \mathbf{R}^m .

3. PROPOSED ALGORITHM

3.1. A projected gradient algorithm

The reduced problem (5) obtained after squeezing is structurally different from (1). Therefore, numerical procedures devised to solve (1) (*e.g.*, [14]) cannot be re-used to address (5). We notice, however, that (5) is a quadratic program with linear constraints and the latter can therefore be solved with standard convex optimization techniques. In [17], we particularize a (rescaled) projected gradient algorithm [18, Section 2.3] to solve (5). Starting from a current estimate $(\mathbf{q}^{(t)}, \tilde{w}^{(t)})$, each iteration consists in repeating the three following steps. First, during the gradient step, the current iterate is updated in the direction of the negative gradient of the cost function. Second, the projection step requires projecting the updated estimate onto the convex set defined by the linear constraints of problem (5). Although this projection does not admit any closed-form solution, we propose in [17] a procedure that computes it exactly in a *finite* and upper-bounded number of steps. Finally, the last step consists in a convex combination of the current update of the iterate and $(\mathbf{q}^{(t)}, \tilde{w}^{(t)})$. The interested reader is referred to [17, Section 4.B] for a fully-detailed explanation of the method.

3.2. Static versus dynamic squeezing

Similarly to screening methods, the proposed safe squeezing procedure can be used either in a ‘‘static’’ or a ‘‘dynamic’’ way. Static squeezing refers to the case where (12) is applied once for all (with $\mathcal{I} = \emptyset$) on the columns of matrix \mathbf{A} before the application of a numerical optimization procedure. In the numerical experiments presented in this paper, we will rather focus on dynamic squeezing where the tests are interleaved with the iterations of the projected gradient algorithm described above. For this reason, our method will be referred to as *squeezing Projected Gradient* (sPG) in what follows. In such a setting, each test leverages the current estimate (at iteration t) to gradually narrow the safe region and (hopefully) increase the number of saturated entries detected by squeezing test (12). Here, since the proposed projected gradient method only produces a couple of primal iterates $(w^{(t)}, \mathbf{q}^{(t)})$ at each iteration, we compute a dual feasible point by dual scaling of $\mathbf{z} = \mathbf{y} - w^{(t)} \mathbf{s} - \mathbf{A}_{\mathcal{I}} \mathbf{q}^{(t)}$ (see (16)).

4. NUMERICAL EXPERIMENTS

In this section, we assess the performance of the squeezing methodology presented in Section 2. For the two experiments, our simulation setup is as follows. For each trial, new realizations of dictionary \mathbf{A} and observation \mathbf{y} are generated. The observation \mathbf{y} is drawn according to a standard normal distribution. The entries of matrix \mathbf{A} are i.i.d. realizations of a uniform distribution on $[0, 1]$. The columns of \mathbf{A} are then normalized such that $\|\mathbf{a}_i\|_2 = 1$ for all i . Note that we only con-

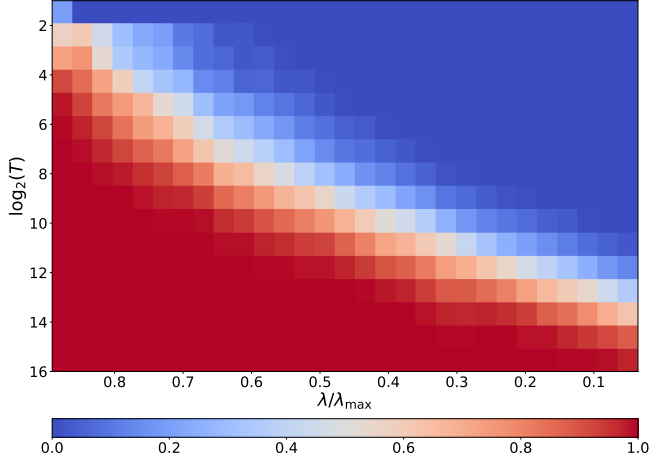


Fig. 1. Proportion of saturated entries as a function of λ/λ_{\max} and the number of iterations T .

sider one family of dictionaries due to space limitation. We refer the reader to [17] for additional experiments. Finally, for the two experiments, we set $(m, n) = (100, 150)$ so that $\mathbf{A} \in \mathbf{R}^{100 \times 150}$.

4.1. Proportion of saturated entries detected

We first explore the effectiveness of the squeezing procedure for several values of the regularization parameter λ . In particular, we are interested in the number of iterations needed to identify all the saturated entries.

Figure 1 presents the proportion of saturated entries that passes the GAP sphere squeezing test (12) as a function of the number of iterations T of the projected gradient method. These results have been obtained by averaging the following setup 20 times. For each couple of realization (\mathbf{A}, \mathbf{y}) , problem (1) is solved $\forall \lambda$ such that $\lambda/\lambda_{\max} \in \{10^{-0.05\ell}\}_{\ell=1}^{31}$ (recall that λ_{\max} has been defined in (3)) with sPG. In particular, the solver is run until a duality gap of 10^{-9} is reached. The resulting “high accuracy solution” is used to identified the total number of saturated entries.

One observes that a large proportion of saturated entries is detected after a few iterations. We note that the proposed GAP sphere squeezing test remains relevant even when the ratio λ/λ_{\max} becomes small.

4.2. Gains in the computation of solution paths

The main interest of safe squeezing is to reduce computation costs. Hence, in this second experiment, we investigate the computational burden needed to compute solutions paths to a prescribed accuracy with (resp. without) safe squeezing. More precisely, the total number of operations² needed to

²We restrict our attention to multiplications since they entail a much higher computational burden than additions in floating-point arithmetic.

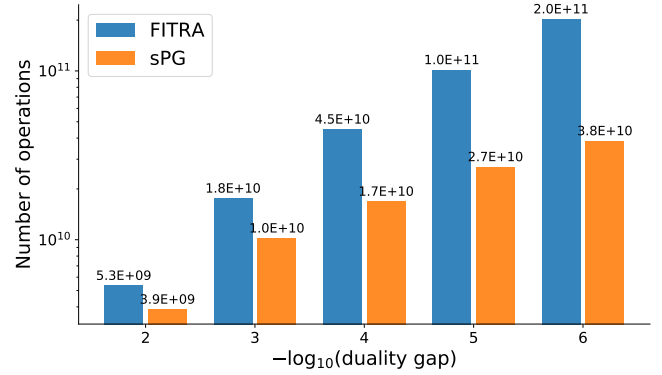


Fig. 2. Number of operations needed to compute a solution path as a function of the duality gap for FITRA [14] and sPG.

compute these paths with different solvers are compared. Depending on the type of implementation (sequential or parallel), such a figure of merit can be interpreted as either computational or energy savings. More precisely, we consider a sequential implementation of antispase coding where the solution of (1) is computed for $\lambda_{\ell}/\lambda_{\max} = 10^{-0.025\ell}$ for $\ell \in \{1, \dots, 60\}$ and up to different accuracy in terms of duality gap. For each λ_{ℓ} , we search for a solution of (1) with each iterative procedures initialized with the solution obtained for $\lambda_{\ell} - 1$ (warm start). Finally, we compare sPG (see Section 3) to FITRA [14] a accelerated proximal gradient algorithm, which is currently the state of the art to solve (1).

Figure 2 illustrates the computational cost of each path seen as a function of the target duality gap. The results are averaged over 20 realizations. One observes that the proposed safe squeezing methodology leads to substantial gains. For example, sPG requires more than 2 times less operations than FITRA as soon as the target duality gap is smaller than 10^{-2} . The gain of the proposed procedure increases for higher accuracies: one can obtain savings of one order of magnitude when the target duality gap is equal to 10^{-6} .

5. CONCLUSIONS

In an effort to speed up the computation of antispase representations in the context of large-scale scenarios, we proposed a new methodology, dubbed *safe squeezing*, to detect saturated entries in the solution of the optimization problem. The proposed procedure enables to transform the initial problem into another (equivalent) problem of smaller dimension, allowing for potentially dramatic complexity reduction. Numerical simulations demonstrated the practical benefits of our methodology. Future work includes the design of squeezing tests based on safe regions having more refined geometries such as dome [19, 20].

References

- [1] Simon Foucart and Holger Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser Basel, 2013.
- [2] Amir Beck and Marc Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Img. Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.
- [3] Laurent El Ghaoui, Vivian Viallon, and Tarek Rabbani, “Safe feature elimination in sparse supervised learning,” Tech. Rep. UC/EECS-2010-126, EECS Dept., University of California at Berkeley, Sep. 2010.
- [4] Olivier Fercoq, Alexandre Gramfort, and Joseph Salmon, “Mind the duality gap: safer rules for the lasso,” in *Proc. Int. Conf. Machine Learning (ICML)*, Lille, France, Jul. 2015, vol. 37 of *Proceedings of Machine Learning Research*, pp. 333–342.
- [5] Antoine Bonnefoy, Valentin Emiya, Liva Ralaivola, and Rémi Gribonval, “Dynamic Screening: Accelerating First-Order Algorithms for the Lasso and Group-Lasso,” *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 20, 2015.
- [6] A. Malti and C. Herzet, “Safe screening tests for lasso based on firmly non-expansiveness,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP)*, Mar. 2016, pp. 4732–4736.
- [7] C. Herzet, C. Dorffer, and A. Drémeau, “Gather and conquer: Region-based strategies to accelerate safe screening tests,” *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3300–3315, Jun. 2019.
- [8] Eugene Ndiaye, Olivier Fercoq, Alexandre Gramfort, and Joseph Salmon, “Gap safe screening rules for sparse multi-task and multi-class models,” in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, Cambridge, MA, USA, 2015, NIPS’15, pp. 811–819, MIT Press.
- [9] Eugene Ndiaye, Olivier Fercoq, Alexandre Gramfort, and Joseph Salmon, “Gap safe screening rules for sparsity enforcing penalties,” *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 4671–4703, Jan. 2017.
- [10] G. Gasso A. Rakotomamonjy and J. Salmon, “Screening rules for lasso with non-convex sparse regularizers,” in *Proc. Int. Conf. Machine Learning (ICML)*, 2019.
- [11] Brendan Farrell and Peter Jung, “A Kashin approach to the capacity of the discrete amplitude constrained Gaussian channel,” in *Proc. Int. Conf. Sampling Theory and Applications (SAMPTA)*, Marseille, France, May 2009.
- [12] J. Ilic and T. Strohmer, “Papr reduction in ofdm using kashin’s representation,” in *2009 IEEE 10th Workshop on Signal Processing Advances in Wireless Communications*, Jun. 2009, pp. 444–448.
- [13] Hervé Jegou, Teddy Furon, and Jean-Jacques Fuchs, “Anti-sparse coding for approximate nearest neighbor search,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP)*, 2012, pp. 2029–2032.
- [14] Christoph Studer and Erik G. Larsson, “PAR-aware large-scale multi-user MIMO-OFDM downlink,” *IEEE J. Sel. Areas Comm.*, vol. 31, no. 2, pp. 303–313, Feb. 2013.
- [15] Clément Elvira, Pierre Chainais, and Nicolas Dobigeon, “Bayesian antisparsity coding,” *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1660–1672, Apr. 2017.
- [16] M. Vural, P. Jung, and S. Stańczak, “A new outlier detection method based on anti-sparse representations,” in *2017 25th Signal Processing and Communications Applications Conference (SIU)*, May 2017, pp. 1–4.
- [17] Clément Elvira and Cédric Herzet, “Safe squeezing for antisparsity coding,” Available at <http://people.rennes.inria.fr/Cedric.Herzet/>, Oct. 2019.
- [18] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 2nd edition, 1999.
- [19] Z. J. Xiang and P. J. Ramadge, “Fast lasso screening tests based on correlations,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP)*, Mar. 2012, pp. 2137–2140.
- [20] Z. J. Xiang, Y. Wang, and P. J. Ramadge, “Screening tests for lasso problems,” *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 39, no. 5, pp. 1008–1027, May 2017.